UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/852,620	05/11/2001	Osamu Hori	208447US2SRD	5332

22850 7590 02/05/2004

OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
1940 DUKE STREET
ALEXANDRIA, VA 22314

EXAMINER	
CUNNINGHAM, GREGORY F	
ART UNIT	**PAPER NUMBER**
2676	10

DATE MAILED: 02/05/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No. 09/852,620	Applicant(s) HORI ET AL.	
	Examiner Greg Cunningham	Art Unit 2676	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *14 November 2003*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-39* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-39* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *11 May 2001* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☒ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☒ All b)☐ Some * c)☐ None of:

1.☒ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____.

DETAILED ACTION

1. This action is responsive to communications of amendment received 11/14/2003.

2. The disposition of the claims is as follows: claims 1-39 are pending in the application. Claims 1, 6, 11, 16, 20, 24 and 28-39 are independent claims.

Information Disclosure Statement

3. The information disclosure statement filed 5/11/2001 fails to comply with 37 CFR 1.98(a)(2), which requires a legible copy of each U.S. and foreign patent; each publication or that portion which caused it to be listed; and all other information or that portion which caused it to be listed. Patent serial number 09/852,620 is listed but not present.

4. Both IDS submissions are missing form PTO-1449, therefore no signed PTO-1449 is enclosed.

Specification

5. In view of amended specification, objection is withdrawn.

Claim Objections

6. Although claim 28 is not identified in applicant's amendment as being (Currently Amended) but rather (Original) - not amended, even though it, claim 28, has actually been amended in accordance with claim objection item number six (6) of prior office action, paper number six (6). Therefore in light of amended claim 28, objection is withdrawn.

Claim Rejections - 35 USC § 112

7. In view of amended claims 6 and 9, 112 rejections are withdrawn.

Claim Rejections - 35 USC § 102

8. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (a) the invention was known or used by others in this country, or patented or described in a printed publication in this or a foreign country, before the invention thereof by the applicant for a patent.

9. Claims 1-10, 28, 29, 34, and 35 are rejected under 35 U.S.C. 102(a) as being disclosed by Lee et al., (US Patent Publication 2001/0048753 A1), hereafter Lee.

A. Claim 1, "A method of describing object region data about an object in video data over a plurality of frames, said method comprising:

approximating trajectories with functions, the trajectories being obtained by arranging, in the frames advancing direction, reference position data about one of said plurality of points in each of said frames and relative position data about remaining points in each of said frames, the relative position data referring to the reference position data in the same frame with reference to said one of said plurality of points; and describing the object region data using the functions" is disclosed by Lee in para. 15, 16, 41-42, 45, 49, 51, 73 and 85 at para. [0015], "Motion estimation techniques, such as global and local motion estimation, are used to track the movement of the object through the video sequence."; in para. "[0016] Preferably, the user is presented with a graphical user interface showing a frame of video data, and the user identifies, with a mouse, pen, tablet, etc., **the rough outline of an object by selecting points around the perimeter of**

the object. Curve-fitting algorithms can be applied to fill in any gaps in the user-selected points. After this initial segmentation of the object, the unsupervised tracking is performed. During unsupervised tracking, the motion of the object is identified from frame to frame. The system automatically locates similar semantic video objects in the remaining frames of the video sequence, and the identified object boundary is adjusted based on the motion transforms.";

in para. [0019] "Thus, a computer can be programmed with software programming instructions for implementing a method of tracking rigid and non-rigid motion of an object across multiple video frames. The object has a perimeter, and initially a user identifies a first boundary approximating this perimeter in a first video frame. A global motion transformation is computed which encodes the movement of the object between the first video frame and a second video frame. The global motion transformation is applied to the first boundary to identify a second boundary approximating the perimeter of the object in the second video frame. By successive application of motion transformations, boundaries for the object can be automatically identified in successive frames.";

in para. [0021] "A motion transformation function representing the transformation between the object in the first frame and the object of the second frame, can be applied to the outline to warp it into a new approximate boundary for the object in the second frame."; "[0041] FIG. 1 shows the two basic steps of the present system of semantic video object extraction. In the first step 100, the system needs a good semantic boundary for the initial frame, which will be used as a starting 2D-template for successive video frames. *'approximating the object using a figure for each of said frames;'* During this step a user indicates 110 the rough boundary of a semantic video object in the first frame with an input device such as a mouse, touch sensitive

surface, pen, drawing tablet, or the like. Using this initial boundary, the system defines one boundary lying inside in the object, called the In boundary 102 and another boundary lying outside the object, called Out boundary 104. These two boundaries roughly indicate the representative pixels inside and outside the user-identified semantic video object. These two boundaries are then snapped 106 into a precise boundary that identifies an extracted semantic video object boundary. Preferably the user is given the opportunity to accept or reject 112, 114 the user selected and computer generated outlines. [0042] The goal of the user assistance is to provide an approximation of the object boundary by just using the input device, without the user having to precisely define or otherwise indicate control points around the image feature. Requiring precise identification of control points is time consuming, as well as limiting the resulting segmentation by the accuracy of the initial pixel definitions. A preferred alternative to such a prior art method is to allow the user to identify and portray the initial object boundary easily and not precisely, and then have this initial approximation modified into a precise boundary.";

in para. [0044] "That is, tracking function 118 is able to compute a new approximate boundary for the semantic object in current frame F.sub.1 by adjusting previous boundary data S.sub.0 according to motion data V.sub.0.";

in para. "[0045] Both steps 100 and step 108 require the snapping of an approximate boundary to a precise one. As described below, a morphological segmentation can be used to refine the initial user-defined boundary (step 110) and the motion compensated boundary (S.sub.0) to get the final precise boundary of the semantic video object.";

in para. [0049] "The second is a contour-based method in which a user only indicates control points *'extracting a plurality of points representing the figure for each of said frames;'* along the outline of an object boundary, and splines or polygons are used to approximate a boundary based upon the control points. *'approximating the object using a figure for each of said frames;'* The addition of Splines is superior over the first method because it allows one to fill in the gaps between the indicated points. The drawback, however, is that a spline or polygon will generally produce a best-fit result for the input points given. With few points, broad curves or shapes will result. Thus, to get an accurate shape, many points need to be accurately placed about the image feature's true boundary. But, if it is assumed n nodes guarantees a desired maximal boundary approximation error of e pixels, at a minimum the user must then enter n keystrokes to define a border. For complex shapes, n may be a very large number. In order to avoid such reduce user effort, n can be decreased, but this approach yields larger e vales."

"[0051] As shown, a user has marked, with white points, portions of the left image 148 to identify an image feature of interest. Although it is preferable that the user define an entire outline around the image feature, doing so is unnecessary. As indicated above, gaps in the outline will be filled in with the hybrid pixel-polygon method. The right image 150 shows the initial object boundary after gaps in the initial outline of the left image 148 have been filled in. By allowing the user to draw the outline, the user is able to define many control points without the tedium of specifying each one individually. *'extracting a plurality of points representing the figure for each of said frames;'* in the prior art, allowing such gaps in the border required a tradeoff between precision and convenience. The present invention avoids such a tradeoff by defining In and Out boundaries and modifying them to precisely locate the

actual boundary of the (roughly) indicated image feature."

"[0073] When there are no more pixels to classify, pixels assigned to a In marker are pixels interior to the image feature (semantic video object) defined by the user (FIG. 1, step 110), and pixels assigned to an Out marker are similarly considered pixels exterior to the semantic object. As with pixel-wise classification, the locations where the In and Out pixel regions meet identifies the semantic object's boundary. The combination of all In pixels constitutes the segmented semantic video object."

"[0085] [Returning to FIG. 8, after prediction 350, the next step is motion estimation 352. It is somewhat axiomatic that a good estimation starts with a good initial setting. By recognizing that in the real world the trajectory of an object is generally smooth, this information can be applied to interpreting recorded data to improve compression efficiency. For simplicity, it is assumed that the trajectory of a semantic video object is basically smooth, and that the motion information in a previous frame provides a good guess basis for motion in a current frame. Therefore, the previous motion parameters can be used as the starting point of the current motion estimation process. (Note, however, that these assumptions are for simplicity, and all embodiments need not have this limitation.) For the first motion estimation, since there is no previous frame from which to extrapolate, the initial transformation is set to a=e=1, and b=c=d=f=g=h=0.]

Wherein "the relative position data" [**perimeter** of the object] refers to "the reference position" [object] in the same frame. And the predetermined frame corresponds to the current frame rather then the next frame.

B. Claim 2, *'The method according to claim 1, wherein said object region data comprises information representing a range of frames in which the object exists in the video data and information identifying the figure approximating the object region.'* is disclosed supra for claim 1 by Lee and in para. 17 at "[0017] Mathematical morphology and global perspective motion estimation/-compensation (or an equivalent object tracking system) is used to accomplish these unsupervised steps. Using a set-theoretical methodology for image analysis (i.e. providing a mathematical framework to define image abstraction), mathematical morphology can estimate many features of the geometrical structure in the video data, and aid image segmentation. Instead of simply segmenting an image into square pixel regions unrelated to frame content (i.e. not semantically based), objects are identified according to a semantic basis and their movement tracked throughout video frames. This object-based information is encoded into the video data stream, and on the receiving end, the object data is used to re-generate the original data, rather than just blindly reconstruct it from compressed pixel regions. Global motion estimation is used to provide a very complete motion description for scene change from frame to frame, and is employed to track object motion during unsupervised processing. However, other motion tracking methods, e.g. block-based, mesh-based, parametric estimation motion estimation, and the like, may also be used."

C. Claim 3, *'The method according to claim 1, wherein said object region data comprises one of information representing related information linking to the object and information representing a method of accessing the related information.'* is disclosed supra for claim 1, particularly in para. 49.

D. Claim 4, *'The method according to claim 1, wherein said relative position data are components of differential vectors between the one of said plurality of points and remaining points.'* is disclosed supra for claim 1 and in para. 81 at "[0081] The algorithm computes the partial derivatives of e.sub.j in the semantic video object with respect to the unknown motion parameters (a, b, c, d, e, f, g). That is, 1 e i m 0 = x i D i I ' x ' e i m 7 = y i D i (x i ' I ' x ' + y i ' I ' y ') a kl = i e i m k e i m l b k = - i e i e i m k

[0082] where D.sub.j is the denominator, I'=F.sub.k', I=F.sub.k-1 and (m.sub.0, m.sub.1, m.sub.2, m.sub.3, m.sub.4, m.sub.5, m.sub.6, m.sub.7)=(a, b, c, d, e, f, g, h)."

E. Claim 5, *'The method according to claim 1, wherein said object region data comprises parameters of the functions.'* is disclosed supra for claim 1 and in para. 86 at "[0086] Once motion prediction 350 and estimation 352 is computed, the previous boundary is then warped 354 according to the predicted motion parameters (a, b, c, d, e, f, g, h), i.e., the semantic object boundary in the previous frame (B.sub.i-1) is warped towards the current frame to become to current estimate boundary (B.sub.i'). Since the warped points generally do not fall on integer pixel coordinates, an inverse warping process is performed in order to get the warped semantic object boundary for the current frame. Although one skilled in the art will recognize that alternate methods may be employed, one method of accomplishing warping is as follows."

F. Claim 6, "A method of describing object region data about an object in video data over a plurality of frames, said method comprising: approximating the object using a figure for each of said frames; extracting a plurality of points representing the figure for each of said frames; approximating trajectories with functions, the trajectories being obtained by arranging, in the frames advancing direction, reference position data about said plurality of points in a

predetermined frame and relative position data about said plurality of points in a succeeding frame, the relative position data referring to the reference position data in the same frame; and describing the object region data using the functions." is disclosed supra for claim 1 and in para. [74-77] at "[0074] FIG. 8 is a flowchart showing automatic subsequent-frame boundary tracking, performed after a semantic video object has been identified in an initial frame, and its approximate boundary adjusted (i.e. after pixel classification). Once the adjusted boundary has been determined, it is tracked into successive predicted frames. Such tracking continues iteratively until the next initial frame '*reference frame*' (if one is provided for). Subsequent frame tracking consists of four steps: motion prediction 350, motion estimation 352, boundary warping 354, and boundary adjustment 356. Motion estimation 352 may track rigid-body as well as non-rigid motion.

[0075] In a given frame sequence, there are generally two types of motion, rigid-body in-place movement and translational movement. Rigid motion can also be used to simulate non-rigid motion by applying rigid-motion analysis to sub-portions of an object, in addition to applying rigid-motion analysis to the overall object. Rigid body motion can be modeled by a perspective motion model. That is, assume two boundary images under consideration are B.sub.k-1(x, y) which includes a boundary indicating the previous semantic video object, and a current boundary indicated by B.sub.k(x', y'). Using the homogeneous coordinates *(coordinate 'reference frame' implied)*, a 2D planar perspective transformation can be described as:

x'=(a*x+b*y+c)/(g*x+h*y+1)

[0076] y'=(d*x+e*y+f)/(g*x+h*y+1)

[0077] The perspective motion model can represent a more general motion than a translational or affine motion model, such that if g=h=0 and a=1, b=0, d=0, e=1, then x'=x+c and y'=y+f, which becomes the translational motion model. Also, if g=h=0, then x'=a*x+b*y+c and y'=d*x+e*y+f, which is the affine motion model."

Wherein "succeeding frame " corresponds to [next frame or next video frame or next one].

G. Per dependent claims 7-10, these are directed to a method for performing the method of dependent claims 2-5, respectively, and therefore are rejected to dependent claims 2-5.

H. Per independent claims 28, 29 and 34, 35, these are directed to a article of manufacture and computer data signal, respectively, for performing the method of independent claims 1 and 6, respectively, and therefore are identically rejected to independent claims 1 and 6.

Claim Rejections - 35 USC § 103

10. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

11. Claims 11-23, 30-32 and 36-38 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lee et al., (US Patent Publication 2001/0048753 A1), as applied to claims 1-5 above, and further in view of Jasinschi et al., (US Patent Number 6,504,569 B1), hereafter Jasinschi.

A. Claim 11, '*A method of describing object region data about an object in video data over a plurality of frames, said method comprising: approximating the object using a figure for each*

of said frames; extracting a plurality of points representing the figure for each of said frames; approximating trajectories with functions, the trajectories being obtained by arranging, in the frames advancing direction, data indicating positions of said plurality of points; and describing the object region data using the functions and depth information of the object.' is disclosed by Lee supra for claim 1. However Lee does not appear to disclose *'describing the object region data using depth information of the object'*, but Jasinschi does in col. 1, lns. 37-58 at "(10) Accordingly the present invention provides a method of generating 2-D extended images from 3-D data extracted from a video sequence representing a natural scene. In an image pre-processing stage image feature points are determined and subsequently tracked from frame to frame of the video sequence. In a structure-from-motion stage the image feature points are used to estimate three-dimensional object velocity and depth. Following these stages depth and motion information are post-processed to generate a dense three-dimensional depth map. World surfaces, corresponding to extended surfaces, are composed by integrating the three-dimensional depth map information.";

in col. 2, lns. 52-65; and col. 3, lns. 33-36.

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to apply object tracking disclosed by Lee in combination with depth determining information disclosed by Jasinschi, and motivated to combine the teachings because it would provide a method of generating 2-D extended images from 3-D data extracted from a two-dimensional video sequence as revealed by Jasinschi in col. 1, lines 32-34.

B. Per dependent claims 12 and 13, these are directed to a method for performing the method of dependent claims 2 and 3, respectively, and therefore are rejected to claim 11 and to dependent claims 2 and 3.

C. Per dependent claim 14, *"The method according to claim 11, wherein said object region data is described by using the depth information of the object and parameters of the functions."* is disclosed supra by Lee for claim 4 and supra by Jasinschi for claim 11.

D. Per dependent claim 15, *"The method according to claim 11, wherein said depth information is a relative depth and has a discrete level value."* is disclosed supra by Lee for claim 4 and supra by Jasinschi for claim 11 and in col. 7, lns. 14-19 at "Step 4: Extract the camera rotation matrix R and the camera translation vector T from the computed essential matrix E. Step 5: Given R and T estimate the depth Z.sub.i at every feature point F.sup.i.sub.k. "

E. Claim 16, *'A method of describing object region data about an object in video data over a plurality of frames, said method comprising: approximating the object using a figure for each of said frames; extracting a plurality of points representing the figure for each of said frames; approximating trajectories with functions, the trajectories being obtained by arranging, in the frames advancing direction, data indicating positions of said plurality of points; and describing the object region data using the functions and display flag information indicating a range of frames in which the object or each of said points is visible or not.'* is disclosed by Lee supra for claim 1. However Lee does not appear to disclose '*display flag information indicating a range of frames in which the object or each of said points is visible or not.*', but Jasinschi does in col. 4, lns. 20-28 at "The inputs to the 3-D camera parameter estimator 16 are raw video images, denoted by I.sub.k, and the corresponding "alpha" images, denoted by A.sub.k. The alpha image

is a binary mask that determines the "valid" regions inside each image, i.e., the regions of interest or objects, as shown in FIG. 3 where FIG. 3A represents an image I.sub.k from a tennis match and FIG. 3B represents the alpha image A.sub.k for the background object with the tennis player blanked out." Wherein [binary mask] corresponds to "display flag information"; and [valid regions] corresponds to "object is visible or not".

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to apply object tracking disclosed by Lee in combination with alpha images A.sub.k disclosed by Jasinschi, and motivated to combine the teachings because it would provide a method of generating 2-D extended images from 3-D data extracted from a two-dimensional video sequence as revealed by Jasinschi in col. 1, lines 32-34.

F. Per dependent claims 17 and 18, these are directed to a method for performing the method of dependent claims 2 and 3, respectively, and therefore are rejected to claim 16 and to dependent claims 2 and 3.

G. Per dependent claim 19, *"The method according to claim 16, wherein said object region data is described by using the display flag information and parameters of the functions."* is disclosed supra by Lee and Jasinschi for claim 16 supra. Wherein alpha images A.sub.k corresponds to display flag information for valid regions.

H. Claim 20, *'A method of describing object region data about an object in video data over a plurality of frames, said method comprising: approximating the object using a figure for each of said frames; extracting a plurality of points representing the figure for each of said frames; approximating trajectories with functions, the trajectories being obtained by arranging, in the frames advancing direction, data indicating positions of said plurality of points; and describing*

the object region data using the functions and object passing range information indicating a range where the figure approximating the object exist over said plurality of frames.' is disclosed by Lee and Jasinschi supra for claims 11 and 16. In particular Lee discloses '*describing the object region data using the functions and object passing range information indicating a range where the figure approximating the object exist over said plurality of frames.*' in para. 90-93 at "Sample Output

[0090] FIGS. 11-13 show sample output from the semantic video object extraction system for several video sequences. These sequences represent different degrees of extraction difficulty in real situations. To parallel the operation of the invention, the samples are broken to parts, the first representing initial frame (user assisted) segmentation results, and the second subsequent frame (automatic) tracking results.

[0091] The three selected color video sequences are all in QCIF format (176.times.144) at 30 Hz. The first Akiyo 450 sequence contains a woman sitting in front of a still background. The motion of the human body is relatively small. However, this motion is a non-rigid body motion because the human body may contain moving and still parts at the same time. The goal is to extract the human body 452 (semantic video object) from the background 454. The second Foreman 456 includes a man 458 talking in front of a building 460. This video data is more complex than Akiyo due to the camera being in motion while the man is talking. The third video sequence is the well-known Mobile-calendar sequence 462. This sequence has a moving ball 464 that is traveling over a complex background 466. This sequence is the most complex since the motion of the ball contains not only translational motion, but also rotational and zooming factors.

[0092] FIG. 11 shows initial frame segmentation results. The first row 468 shows an initial boundary obtained by user assistance; this outline indicates an image feature within the video frame of semantic interest to the user. The second row 470 shows the In and Out boundaries defined inside and outside of the semantic video object. For the output shown, the invention was configured with a size of 2 for the square structure element used for dilation and erosion. The third row 472 shows the precise boundaries 474 located using the morphological segmentation tool (see FIG. 6 above). The forth row 476 shows the final extracted semantic objects.

[0093] FIG. 12 shows subsequent frame boundary tracking results. For the output shown, the tracking was done at 30 Hz (no skipped frames). Each column 478, 480, 482 represents four frames randomly chosen from each video sequence. FIG. 13 shows the corresponding final extracted semantic video objects from the FIG. 12 frames. As shown, the initial precise boundary 474 has been iteratively warped (FIG. 8, step 354) into a tracked 484 boundary throughout the video sequences; this allows implementations of the invention to automatically extract user-identified image features."

However Lee does not appear to disclose "describing the object region data using the functions and object passing range information indicating a range where the figure approximating the object exist over said plurality of frames.", but Jasinschi does in col. 4, lns. 20-28 at "The inputs to the 3-D camera parameter estimator 16 are raw video images, denoted by I.sub.k, and the corresponding "alpha" images, denoted by A.sub.k. The alpha image is a binary mask that determines the "valid" regions inside each image, i.e., the regions of interest or objects, as shown in FIG. 3 where FIG. 3A represents an image I.sub.k from a tennis match and FIG. 3B represents the alpha image A.sub.k for the background object with the tennis player blanked

out." Wherein [object passing range information] corresponds to "sub.k"; and [valid regions] corresponds to "object exist".

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to apply object and passing range tracking disclosed by Lee in combination ranging information disclosed by Jasinschi, and motivated to combine the teachings because it would provide a method of generating 2-D extended images from 3-D data extracted from a two-dimensional video sequence as revealed by Jasinschi in col. 1, lines 32-34.

I. Per dependent claims 21 and,22, these are directed to a method for performing the method of dependent claims 2 and 3, respectively, and therefore are rejected to claim 20 and to dependent claims 2 and 3.

J. Per dependent claim 23, *"The method according to claim 20, wherein said object region data is described by using the object passing range information and parameters of the functions."* is disclosed supra by Lee and Jasinschi for claim 20 supra and exemplified by Lee.

K. Per independent claims 30-32 and 36-38, these are directed to a article of manufacture and computer data signal, respectively, for performing the method of independent claims 11, 16, and 20, respectively, and therefore are identically rejected to independent claims 11, 16, and 20.

12. Claims 24-27, 33 and 39 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lee et al., (US Patent Publication 2001/0048753 A1), as applied to claim1-5 above, and further in view of "Panoramic Image Mosaics", Heung-Yeung Shum, hereafter Shum.

A. Claim 24, *'A method of describing object region data about an object moving in a panorama image formed by combining a plurality of frames with being overlapped, said method comprising: approximating the object in the panorama image using a figure; extracting a*

plurality of points representing the figure in a coordinate system of the panorama image; approximating trajectories with functions, the trajectories being obtained by arranging, in the frames advancing direction, data indicating positions of said plurality of points; and describing the object region data using the functions.' is disclosed by Lee supra for claim 1. However Lee does not disclose '*panorama image formed by combining a plurality of frames with being overlapped, said method comprising: approximating the object in the panorama image using a figure; extracting a plurality of points representing the figure in a coordinate system of the panorama image'*, but Shum does in abstract and last paragraph of p. 2, at "This paper presents some techniques for constructing panoramic image mosaics from sequences of images. Our mosaic representation associates a transformation matrix with each input image, rather than explicitly projecting all of the images onto a common surface (e.g., a cylinder). In particular, to construct a full view panorama, we introduce a rotational mosaic representation that associates a rotation matrix (and optionally a focal length) with each input image. A patch-based alignment algorithm is developed to quickly align two images given motion models. Techniques for estimating and refining camera focal lengths are also presented.

In order to reduce accumulated registration errors, we apply global alignment (block adjustment) to the whole sequence of images, which results in an optimally registered image mosaic. To compensate for small amounts of motion parallax introduced by translations of the camera and other unmodeled distortions, we develop a local alignment (deghosting) technique which warps each image based on the results of pairwise local image registrations. By combining both global and local alignment, we significantly improve the quality of our image mosaics, thereby enabling the creation of full view panoramic mosaics with hand-held cameras.

We also present an inverse texture mapping algorithm for efficiently extracting environment maps from our panoramic image mosaics. By mapping the mosaic onto an arbitrary texture-mapped polyhedron surrounding the origin, we can explore the virtual environment using standard 3D graphics viewers and hardware without requiring special-purpose players. Third, any deviations from the pure parallax-free motion model or ideal pinhole (projective) camera model may result in local misregistrations, which are visible as a loss of detail or multiple images (ghosting). To overcome this problem, we compute local motion estimates (block-based optical flow) between pairs of overlapping images, and use these estimates to warp each input image so as to reduce the misregistration.

Therefore it would have been obvious to one of ordinary skill in the art at the time the invention was made to apply object tracking disclosed by Lee in combination with panorama image mosaics disclosed by Shum, and motivated to combine the teachings because it would provide a technique for constructing panoramic image mosaics from sequences of images as disclosed by Shum in abstract.

B. Per dependent claims 25-27, these are directed to a method for performing the method of dependent claims 2, 3, and 5, respectively, and therefore are rejected to claim 24 and to dependent claims 2, 3, and 5.

C. Per independent claim 33 and 39, these are directed to a article of manufacture and computer data signal, respectively, for performing the method of independent claim 24 and therefore are identically rejected to independent claim 24.

Response to Arguments

13. With regard to claims 1 and 6, Lee et al. in fact more than merely teaches that the previous motion parameters can be used as the starting point of the current motion estimation process. In para. [0085] on p. 8, Lee et al. furthermore teach "(Note, however, that these assumptions are for simplicity, and all embodiments need not have this limitation.) For the first motion estimation, since **there is no previous frame from which to extrapolate**, the **initial transformation is set to a=e=1, and b=c=d=f=g=h=0**."

Also in para. [0015], [**Motion estimation techniques**, such as **global** and **local motion estimation**, are used to **track the movement of the object through the video sequence**.] corresponds to "approximating trajectories with functions" and [track the movement of the object through the video sequence] corresponds to "the trajectories being obtained by arranging, in the frames advancing direction" ; in para. [0016], [selecting points] corresponds to "plurality of points"; [**perimeter** of the object] corresponds to "relative position data"; and [object] corresponds to reference position";

in para. [0019] "Thus, a computer can be programmed with software programming instructions for implementing a method of **tracking rigid and non-rigid motion of an object across multiple video frames**. The object has a perimeter, and initially a user identifies a first boundary approximating this perimeter in a first video frame. A global motion transformation is computed which encodes the movement of the object between the first video frame and a second video frame. The global motion transformation is applied to the first boundary to identify a second boundary approximating the perimeter of the object in the second video frame. By successive application of motion transformations, boundaries for the object can be automatically

identified in successive frames."; in para. [0021] "A motion transformation function representing the transformation between the object in the first frame and the object of the second frame, can be applied to the outline to warp it into a new approximate boundary for the object in the second frame."; and in para. [0044] "That is, **tracking function 118** is able to compute a new approximate boundary for the semantic object in current frame F.sub.1 by adjusting previous boundary data S.sub.0 according to motion data V.sub.0."

Wherein "the relative position data" [**perimeter** of the object] refers to "the reference position" [object] in the same frame. And the predetermined frame corresponds to the current frame rather then the next frame.

Wherein "succeeding frame " corresponds to [next frame or next video frame or next one] – claim 6.

With regard to claim 11, [feature points] taught by Jasinschi et al. correspond to "plurality of points" that represent the [object] "figure". Also in col. 2, lns. 52-65; and col. 3, lns. 33-36, [the depth is estimated at the feature points and the resulting values are interpolated/extrapolated to generate a depth value for each image pixel in the scene shot to provide a dense depth map.]

With regard to claim 16, [binary mask] corresponds to flag information; and [valid regions] corresponds to "object is visible or not".

With regard to claim 20, wherein [object passing range information] corresponds to "sub.k"; and [valid regions] corresponds to "object exist".

With regard to claim 24, wherein the [rotational mosaic] corresponds to "moving object", and [rotation matrix] corresponds to "trajectories with functions".

Conclusion

14. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Responses

15. Responses to this action should be mailed to: Commissioner of Patents and Trademarks, Washington, D.C. 20231. If applicant desires to fax a response, (703) 872-9314 may be used for formal communications.

Please label "PROPOSED" or "DRAFT" for informal facsimile communications. Hand-delivered responses should be brought to Crystal Park II, 2121 Crystal Drive, Arlington, VA., Sixth Floor (Receptionist).

Inquiries

16. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Greg Cunningham whose telephone number is (703) 308-6109.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Matthew Bella, can be reached on (703) 308-6829.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703) 305-4700.

G.F. Cunningham

gfc

February 2, 2004

Matthew C. Bella

MATTHEW C. BELLA
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2600